

# A 176x144 148dB adaptive tone-mapping imager

S. Vargas-Sierra, G. Liñán-Cembrano, A. Rodríguez-Vázquez

Instituto de Microelectrónica de Sevilla (IMSE-CNM), CSIC and Universidad de Sevilla  
Avda. Américo Vespucio s/n, 41092 Sevilla, Spain

E-mail: sonia@imse-cnm.csic.es, linan@imse-cnm.csic.es, angel@imse-cnm.csic.es

## ABSTRACT

This paper presents a 176x144 (QCIF) HDR image sensor where visual information is simultaneously captured and adaptively compressed by means of an in-pixel tone mapping scheme.<sup>1</sup> The tone mapping curve (TMC) is calculated from the histogram of a Time Stamp image captured in the previous frame, which serves as a probability indicator of the distribution of illuminations within the present frame. The chip produces 7-bit/pixel images that can map illuminations from 31  $\mu$ lux to 55.3 klux in a single frame in a way that each pixel decides when to stop observing photocurrent integration –with extreme values captured at 8s and 2.34 $\mu$ s respectively. Pixels size is 33x33 $\mu$ m<sup>2</sup>, which includes a 3x3 $\mu$ m<sup>2</sup> Nwell-Substrate photodiode and an autozeroing technique for establishing the reset voltage, which cancels most of the offset contributions created by the analog processing circuitry. Dark signal ( $10.8 \frac{mV}{s}$ ) effects in the final image are attenuated by an automatic programming of the DAC top voltage. Measured characteristics are Sensitivity  $5.79 \frac{V}{lux \cdot s}$ , FWC 12.2ke<sup>-</sup>, Conversion Factor 129(e<sup>-</sup>/DN), and Read Noise 25e<sup>-</sup>. The chip has been designed in the 0.35 $\mu$ m OPTO technology from Austriamicrosystems (AMS). Due to the focal plane operation, this architecture is especially well suited to be implemented in a 3D (vertical stacking) technology using per-pixel TSVs.

**Keywords:** CMOS Image Sensor, High Dynamic Range, Tone Mapping, Focal Plane Processing

## 1. INTRODUCTION

High Dynamic Range (HDR) is required in applications like automotive, surveillance, scientific experiments, etc. In order to solve this issue, High Dynamic Range (HDR) image sensors usually codify illuminations in the scene non-adaptively, using either long bit-words per pixel (e.g. mantissa exponent<sup>2</sup>) obtained from the combination of images captured at different exposures,<sup>3</sup> or a fixed compressive function (e.g. logarithmic approach<sup>4</sup>) among many other possibilities.<sup>5</sup> These non-adaptive approaches usually lead, non-exclusively, to either high computational costs for the post-processing of the images (in the long bit-words case) or the loss of details and lack of contrast (e.g. logarithmic sensors) due to the fixed compression. In order to avoid these drawbacks, the proposed system produces an adaptive compression of illuminations by using only 7-bit per pixel.

## 2. TONE MAPPING ALGORITHM

Tone Mapping techniques are used to compress the information of HDR scenes in Low Dynamic Range (LDR) output representations while preserving the details contained in the scene. These techniques usually belong to the computer graphics field, where the input data is an image with long bit-word per pixel, due to the high computational effort that is required. Tone Mapping algorithms are usually divided depending on the operator applied to the input HDR representation. The main subdivision of Tone Mapping Operators (TMO) are: Global and Local. Global TMO apply the same function to the whole array, named Tone Mapping Curve (TMC). Local TMO apply a different function depending on the pixel.

In order to achieve simultaneous capture and in-pixel compression, a global tone mapping algorithm with very reduced computational effort has been developed taking advantage of the commonly available focal-plane processing circuitry in integration pixels. Moreover, it is adaptive, therefore the global TMC will change in every frame depending on the probability of illuminations in order to optimize the use of the final LDR representation.

The in-pixel operation of the algorithm is based in capturing digital data when the intersection time between the pixel discharge signal  $V_{ph}$  and a global analog reference  $V_{ref}$  occurs. This analog reference  $V_{ref}$  will be fixed during most of the exposition time and will ramp up as fast as possible at the end, which allows poorly illuminated pixels to intersect. In the intersection event, the current status of two digital global references are stored in-pixel: the 4-bit Time Stamp Code

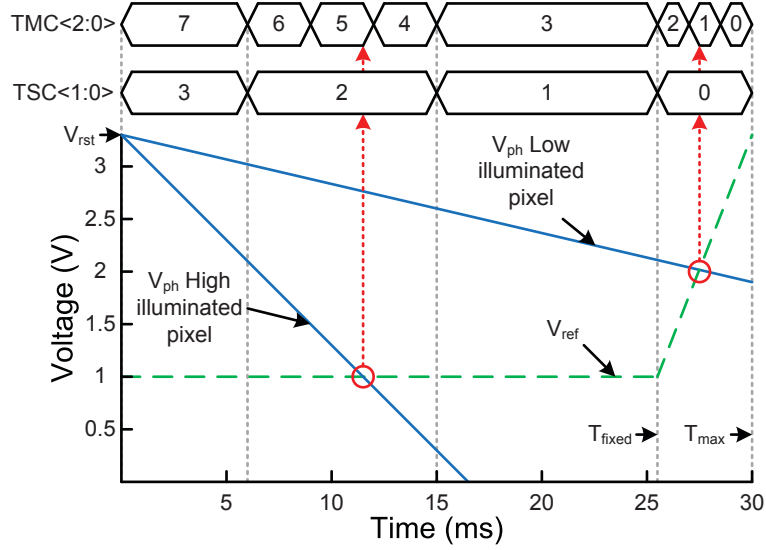


Figure 1: In-Pixel Operation.

(TSC) and the 7-bit Tone Mapping Code (TMC). TSC data is an auxiliary subsampled image used to calculate a histogram, which serves as a illumination probability indicator. TMC defines the global tone mapping function whose non-linear temporal evolution is calculated from our tone-mapping algorithm<sup>1,6</sup> using the probability extracted from the TSC data of the previous frame. The signals involved in the in-pixel operation are shown in Fig. 1, where the example has been simplified with TSC and TMC with only 2-bits and 3-bits, respectively.

This operation performs two kind of compression: the Analog and the Digital compression. The analog compression is derived by the intersection of the discharge pixel signal and the analog reference. Two ratio of analog compression are applied depending on if the pixel signal crosses during the fixed or the ramp-up analog reference. The relation of the crossing time with the current (photocurrent + dark current) that discharges the pixel node  $V_{ph}$  in these regimes are expressed in Eq. 1 for fixed reference and Eq. 2 for voltage ramp-up reference:

$$T_{cross_{fixed}}(I_{pix}, V_{ref}) = \frac{C_{ph}}{I_{pix}}(V_{rst} - V_{fixed}) \quad (1)$$

$$T_{cross_{ramp}}(I_{pix}, V_{ref}) = \frac{C_{ph}}{m \cdot C_{ph} + I_{pix}}(V_{rst} - V_{fixed} + m \cdot T_{fixed}) \quad (2)$$

where  $C_{ph}$  is the capacitance of the integration node,  $I_{pix}$  is the current that discharges the node,  $V_{rst}$  is the reset voltage that is the starting point of discharge,  $V_{fixed}$  is the value of the analog reference during the time that it is constant,  $m$  is the slope of the ramp and  $T_{fixed}$  is the time when the fixed voltage ends and starts the ramp. It must be observed that, the limit  $m \rightarrow \infty$  of  $T_{cross_{ramp}}$  is  $T_{fixed}$ . Therefore, the higher the slope the closer is the analog compression of the ramp to a linear behavior as indicated in Eq. 3, which is the equation of the discharge signal of integration pixels. This behavior is typically used in non-HDR image sensors, which perform a linear acquisition of the illumination of the scene, by using single slope ADCs.

$$V_{ph} = V_{rst} - \frac{I_{pix}}{C_{ph}} \cdot T_{fixed} \quad (3)$$

Therefore, once the analog reference is defined, for a given exposition time, the analog compression is also determined. Then, the digital compression is applied to the crossing time by means of the TMC digital reference, which will alter the analog compression to expand the digital codes of high populated illuminations and compress low populated ones.

In order to generate TMC, first, the data of TSC is needed. The TSC<3:0> digital reference is generated dividing the exposure time into 16 non-linearly distributed windows, each of them having a different TS value. The duration of the first 15 temporal windows have been selected so that they are compressed towards the higher illumination bands (shorter intersection times), mimicking the natural compression ( $1/I_{pix}$ ) of the intersection time expressed in Eq. 1. The last and 16<sup>th</sup> temporal window is determined by the ramp duration, which can be as fast as 153.6 $\mu$ s. These temporal windows are named “bins”, as the TSC data is used to perform a histogram.

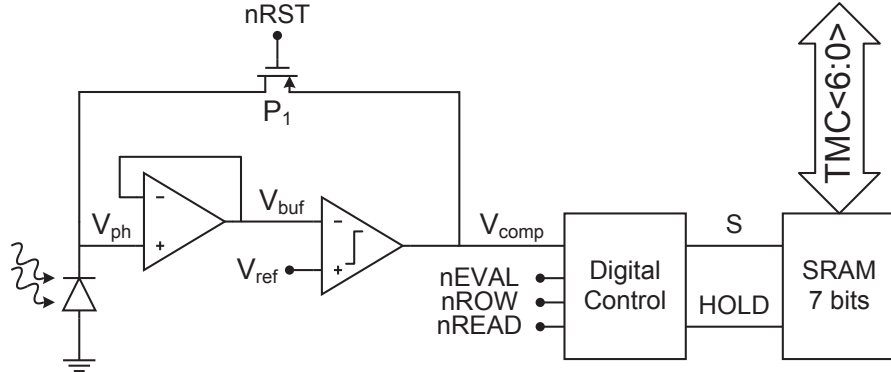
The temporal evolution of TMC<6:0> is created from the histogram of the TS image captured in the previous frame. Due to this fact, we consider the information of TSC as an indicator of probability rather than as an exact evaluation, and so it may fail when the exposure time is too long as compared to the rate of changes in the image. TMC<6:0> varies linearly within each temporal window, spanning over a number of LSBs which is proportional to the weight of this temporal window in the histogram of the TS image. The number LSB spanning during a temporal window is called “levels per bin”. Just for illustration purposes, if the histogram shows that half of the pixels crossed  $V_{ref}$  during temporal window TSC<3:0>=3, the TMC<6:0> curve will span over 64 codes during this temporal window. It is to say that, the levels per bin, for bin=3, is 64. Finally, since the duration of temporal windows are non-linearly distributed in time, the obtained profile for the TMC curve is non-linear in time as well (or piece-wise linear to be more precise). The TMC data will start at the maximum value to decrease non-linearly during the exposition time till the minimum value. This will perform a codification of the higher illuminations with high TMC codes and lower illuminations with low TMC codes, as it is usual in image representation standards.

Once the levels per bin have been obtained, the TMC evolution will be generated decreasing one code any time the data retrieved from a Look Up Table (LUT) indicates it. Each of the 129 position of this LUT (0-128 levels per bin) contains 128 1-bit word which denotes the positions of code decrements in the 128 evaluation cycles of a bin depending of the assigned levels per bin. During every bin period,  $2^7$  bits = 128 evaluations cycles are executed to check whether the pixel signals have crossed the analog voltage reference, in order to store TSC and TMC data. It is worth to mention that within a bin, TMC codes are uniformly distributed in time according to the duration of that bin. Code decrement positions are calculated by the computation of  $\frac{i-128}{N}$  where  $N$  is the levels per bin and  $i = (1, N)$ . These positions are then flipped left-to-right to always obtain a code decrement in the first position of all levels except for position 0, which corresponds to the case that no codes are assigned to a bin, this allows to distinguish the codes between bins. Moreover, the first code decrement order needs to be discarded as it is not necessary.

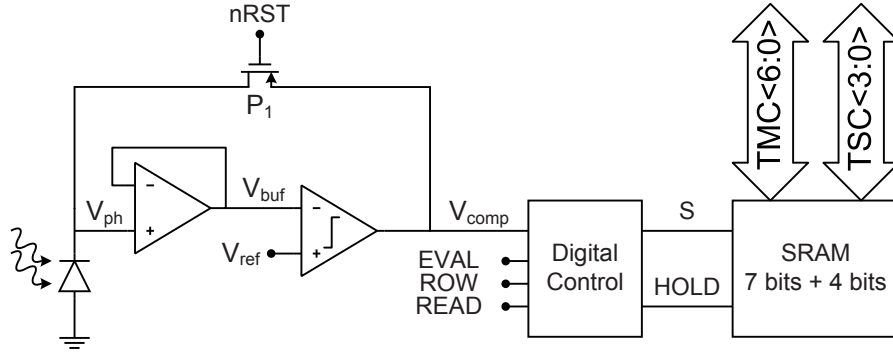
The general algorithm steps to be executed are:

1. Define the duration of the 16 bins (temporal windows)  $\Rightarrow$  define TSC reference.
2. Perform a first acquisition  $\Rightarrow$  the TMC data (final image) can be discarded, as the role for the first frame is storing the TSC data to perform the first histogram.
3. Accumulate TSC data to obtain the histogram.
4. Calculate the levels per bin  $\Rightarrow$  dividing the 16 values that compose the histogram by  $\frac{\text{Number of pixels with TSC data}}{\text{Total number of levels to be assigned}} = \frac{(176-144)/4}{128} = 49.5$ .
5. Floor round results in order to have an integer number of levels per bin, which avoid distributing more than the available 128 TMC codes (levels).
6. The unassigned levels are distributed among the bins according to the higher remainders of the previous division by 49.5.
7. During exposition time, in order to create a piece-wise linear distribution of TMC codes, in every evaluation ( $128 \text{ evaluation} \cdot 16 \text{ bins} = 2048 \text{ evaluations}$ ) a 1-bit word of the LUT is retrieved which indicates if one TMC code decrement must take place or not. This 1-bit word is retrieved depending on levels per bin and index of the evaluation inside the bin, 1 to 128.
8. Download TSC and TMC data  $\Rightarrow$  TMC data is the final image and TSC will be used in step 3 to continue the process.

More details about the algorithm is provided in reference “High-dynamic range tone-mapping algorithm for focal plane processors”.<sup>1</sup>



(a) Basic Pixel Schematic.



(b) Time Stamp Pixel Schematic.

Figure 2: Block Diagrams of the BP and TS Pixels.

### 3. PIXELS

Pixels have been arranged in two categories:

- Basic Pixels: including only TMC sampling circuitry.
- Time Stamp (TS) Pixels: including both TMC and TSC sampling circuitry.

The block level schematic of both pixels are shown in Fig. 2. The sensor, a  $3 \times 3 \mu\text{m}^2$  Nwell/Psubs diode\*, works in photocurrent integration mode. It uses an auto-zeroing technique to establish the reset voltage through the combined action of a buffer (which in operation isolates the photodiode capacitor from comparator's kickback noise), an analog comparator (where  $V_{ref}=V_{rst}$  during reset phase) and a PMOS feedback switch  $P_1$ . Additionally, digital circuitry is included to control read and write operations of the SRAM cells. Signal ROW controls the external write of data row by row (for evaluation and initialization purposes), signal EVAL activates internal write operation and signal READ enables external readouts, which are obviously synchronized with the ROW signal. TS pixels contain  $7(\text{TMC})+4(\text{TSC})=11$  bits of SRAM, whereas Basic Pixels (BP) do only include  $7(\text{TMC})$  SRAM modules.

Pixels are physically arranged as shown in Fig. 3(a). Notice that each TS pixel takes, conceptually, some area from its 3 BP neighbors which is used to allocate the 4 SRAM modules for TSC storage. Indeed, all pixels have  $8(7\text{TMC}+1\text{TSC})$  SRAM blocks. TSC modules are grouped in the middle of the  $2 \times 2$  arrangement, as shown in Fig. 3(a), and controlled by signals produced in the TS pixel only. The layout of a group of  $2 \times 2$  pixels is shown in Fig. 3(b). Observe that we have grouped the SRAM modules in the central vertical region, sharing global control, digital power and ground lines. This increases the attainable pitch and reduces the noise from digital switching in the analog blocks.

\*Aperture in metal structures over the diode is  $9.75 \times 7.30 \mu\text{m}^2$ . Due to this, carriers created within this area can also contribute to the photogenerated current by reaching the photodiode through diffusion, increasing the effective fill-factor.

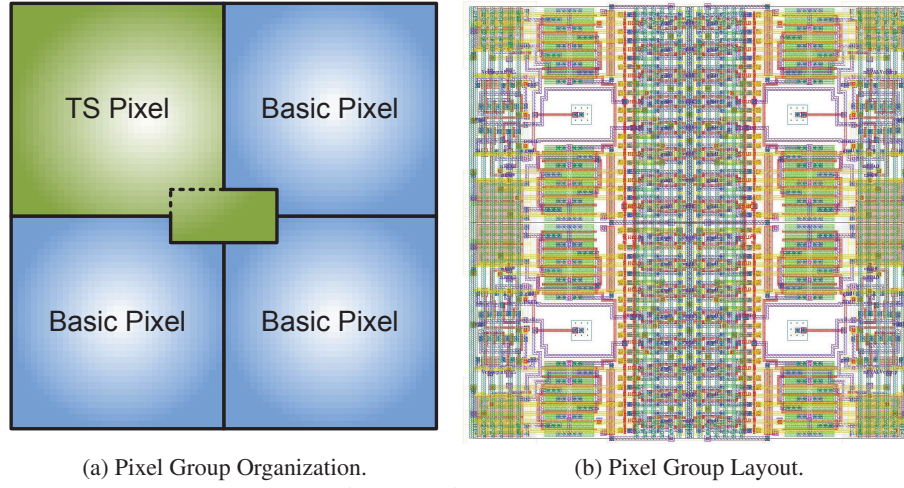


Figure 3: Pixels Group.

### 3.1 Auto-zeroing Technique

A crucial issue in the operation of the imager is the usage of an auto-zeroing technique to cancel out most offset contributions from the two amplifiers in the pixel. During reset phase, the voltage  $V_{rst}$  is applied to the  $V_{ref}$  input in Fig. 2(a,b), and transmitted to the photodiode's integrating capacitor through the negative feedback loop created by the two amplifiers and the reset switch. If we consider that amplifiers can be efficiently modeled to this purpose by their input-referred offset voltage  $V_{ox}$  and a finite DC gain  $A_x$ —where  $x = b$  for the Buffer and  $c$  for the Comparator—, one finds after simple calculations—considering high gains—that the reset value is approximately established to:

$$V_{phrst} = \frac{A_c \cdot (V_{rst} - \frac{A_b}{1+A_b} \cdot V_{ob} + V_{oc})}{1 + \frac{A_c \cdot A_b}{1+A_b}} \approx V_{rst} - V_{ob} + V_{oc} \quad (4)$$

During the integration period (exposure), the feedback switch  $P_1$  is OFF and the photocurrent discharges the integration node from this previous reset voltage. If we consider that turning off the feedback switch introduces a feedthrough error  $\Delta V_{ft}$  in the integration node, we can express the temporal evolution of this node as:

$$V_{ph}(t) = V_{phrst} - \frac{I_{pix}}{C_{pix}} \cdot \Delta t + \Delta V_{ft} \approx V_{rst} - \frac{I_{pix}}{C_{pix}} \cdot \Delta t - V_{ob} + V_{oc} + \Delta V_{ft} \quad (5)$$

If the effective differential input of the comparator is observed, considering very large gains in buffer and comparator and the approximate expression in Eq. 5, it does not contain any reference to neither the comparator nor the buffer input-referred offset voltages. Needless to say, this ideal behavior will not occur in practice, where the output will obviously exhibit some dependency on the offset of these two amplifiers. The residue of the autozero operation at the effective differential input voltage of the comparator is found to be given by:

$$V_{residue} = \frac{A_b + 1}{A_c A_b + A_b + 1} V_{rst} - \frac{1}{A_b + 1} \frac{I_{pix}}{C_{pix}} \cdot \Delta t - \frac{A_b}{A_c A_b + A_b + 1} V_{ob} + \frac{A_b + 1}{A_c A_b + A_b + 1} V_{oc} - \frac{A_b}{A_b + 1} \Delta V_{ft} \quad (6)$$

$$V_{residue} \approx \frac{1}{A_c} V_{rst} - \frac{1}{A_b} \frac{I_{pix}}{C_{pix}} \cdot \Delta t - \frac{1}{A_c} V_{ob} + \frac{1}{A_c} V_{oc} - \Delta V_{ft} \quad (7)$$

Clearly, most of errors (except the feedthrough, which is the main error contribution at the end) vanish as the amplifiers gain is sufficiently high. This, in practice, is translated into a small residual contribution due to the impossibility of designing very large gain low-power amplifiers (each amplifier consumes 50nA) within such small area. The pixel design has been made under the 3 sigma constraint for all added non-idealities.





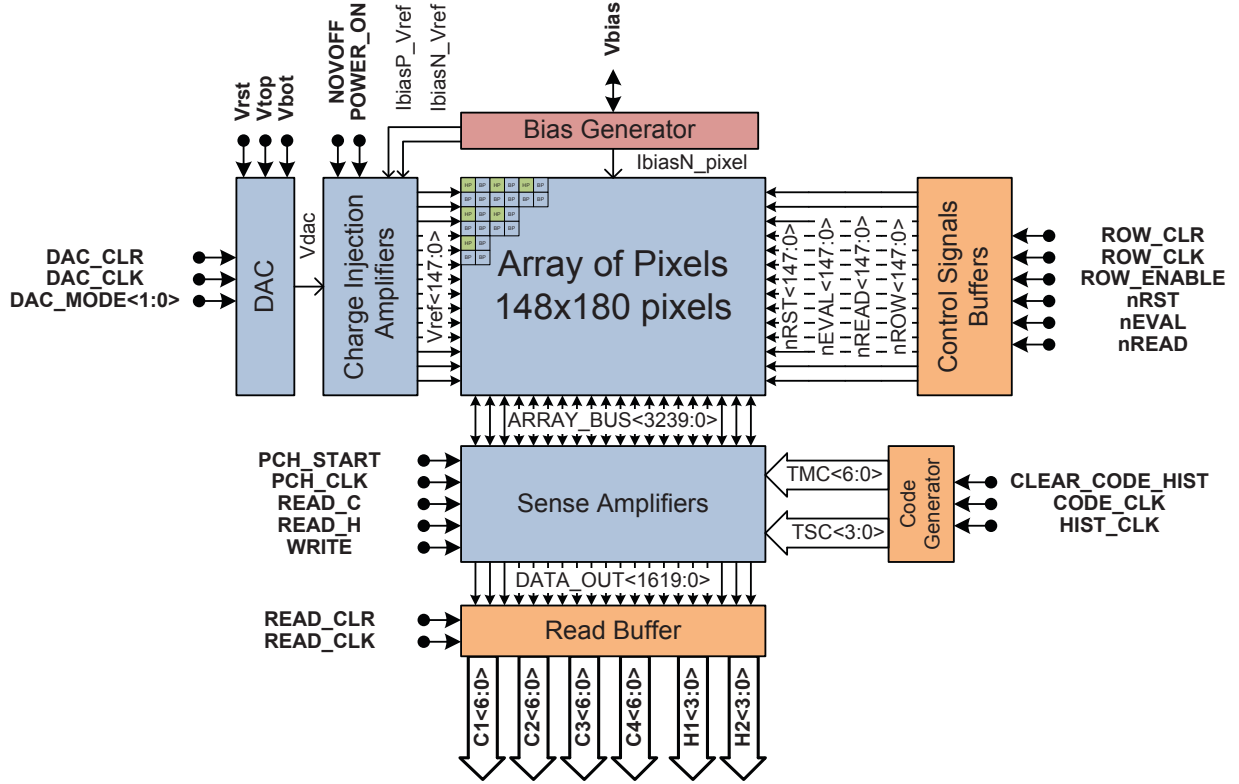


Figure 6: TVHC Chip Block Diagram

## 5. CHIP ARCHITECTURE

The architecture of the chip is shown in Fig. 6, with its core array of  $148 \times 180$  pixels (QCIF + 2 dummy rows and columns on each side). Pixels functionality is supported by additional periphery blocks. An 8-bit DAC generates the reset voltage  $V_{rst}$  during reset, the fixed voltage  $V_{bot}$  during the exposition time and finally the 128 levels ramp signal from  $V_{bot}$  to  $V_{top}$  during the last temporal window. 148 buffers (one per row) enhance the dynamics of distributing  $V_{ref}$  to the array. Digital control signals also employ per-row distributed digital buffers (including clock-tree generation). TSC<3:0> and TMC<6:0> are generated by a Code Generator in gray format. This coding reduces switching at the pixel level to only one SRAM module at a time (instead of 7) for Basic Pixels and 2 (instead of 11) for TS Pixels. Read and write operations from the array are accomplished by a bank of sense amplifiers. Image is retrieved row by row and stored in a read buffer (1 row) which outputs images through a high-speed 36-bit bus (4 TMC codes + 2 TSC codes at a time –eq. to 43MBytes/s). Fig. 7 shows a microscope's capture.

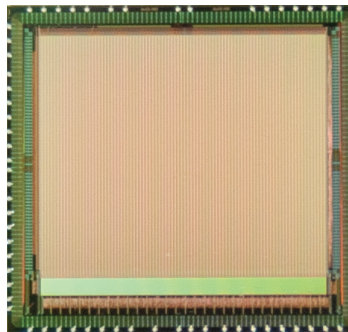


Figure 7: TVHC Microscope Capture

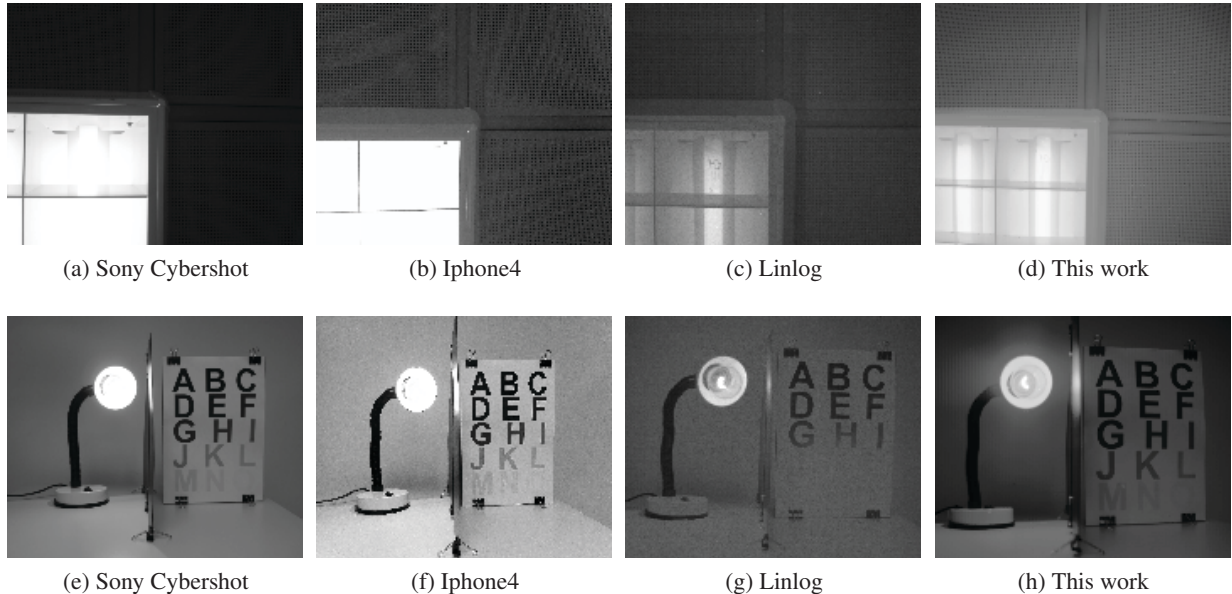


Figure 8: Comparison with commercial cameras.

It must be remarked that, for flexibility and being the first prototype of the idea, the calculations of levels per bin and the LUT have been implemented in an external FPGA. However, the code has been developed, in Verilog, in a way that can be easily implemented by automatic digital synthesis to conform a whole System-On-Chip (SoC) in a future evolution.

## 6. EXPERIMENTAL RESULTS

We present here a comparison of images captured from 3 commercial systems and our chip (see Fig. 8). The Sony Cybershot DSC-W80<sup>7</sup> –which includes an enhanced sensitivity CCD sensor (*Super HAD<sup>TM</sup> CCD*), the Iphone4 camera –which allows HDR Mode<sup>8</sup> (since iOS 4.2) by using a combination of 3 pictures, and the Photonfocus MV-D752E-40-U2-12,<sup>9</sup> which employs the Lin-Log technology. Noticeably, despite using only half of the codes (128 vs. 256) for image representation, our approach produces an image which is –visually– competitive with the other approaches. The Lin-Log sensor shows little more details within lamp areas at the expense of a higher noise. The DSC-W80 produces much lower noise but it shows both over and under exposed areas. Finally, the HDR mode in the Iphone4, shows some similar performance in the darker areas but fails to produce details in the brighter ones. Table 1 summarizes the most important characteristics of the chip. It could be observed comparing the DR (SNR1) of a linear acquisition (obtained using only the intersection times with the ramp) and the DR (SNR1) (obtained with this method) that the DR increment is about 114.2dB.

Table 1: Chip Characteristics

Characteristic	Value	Characteristic	Value
Technology	3.3V 0.35 $\mu$ m 2P4M AMS OPTO	Read Noise	25e <sup>-</sup>
Image Size	180(H)x148(V) (QCIF+dummies)	Sensitivity	5.79V $\cdot$ lux <sup>-1</sup> $\cdot$ s <sup>-1</sup>
Pitch	33 $\mu$ m	Average Dark Signal	10.8mV $\cdot$ s <sup>-1</sup>
Photodiode	NW-Psub, Aperture 9.75 $\times$ 7.3 $\mu$ m <sup>2</sup>	Maximum Dynamic Range (SNR10)	148dB
Fill Factor	0.8% (Diode) 6.5% (Aperture)	Maximum Dynamic Range (SNR1)	168dB
Exposure Time Range	2.34 $\mu$ s to 8s	Linear Dynamic Range (SNR1)	53.8dB
Image coding	7 bits	Fastest Image Download Time	666 $\mu$ s
Full Well Capacity	12.2ke <sup>-</sup>	Maximum Frame Rate	6112 fps
Conversion Factor	129(e <sup>-</sup> /DN)	Fastest Operation Power Consumption	562mW@6112fps



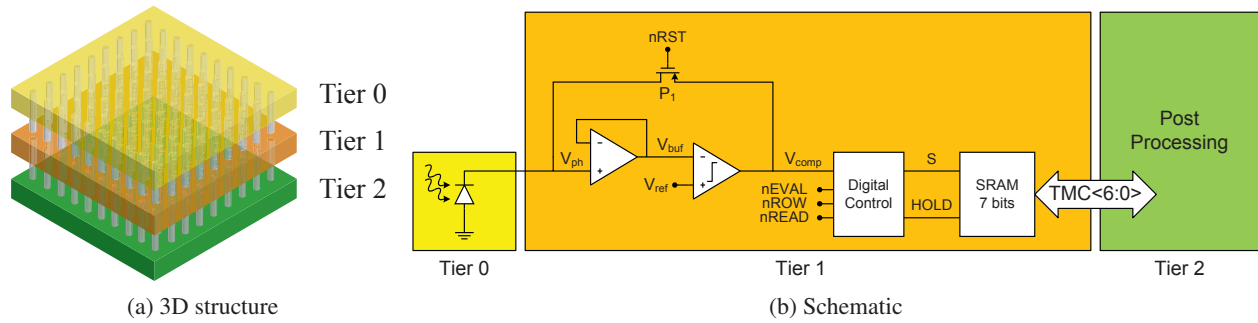


Figure 9: 3D IC future work.

## 7. CONCLUSIONS AND FUTURE WORK

We have presented an imager that automatically adapts to compress the HDR scene in a 7-bit format using a Tone-Mapping algorithm with information from the previous frame. Pixels include auto-zeroing and in-pixel SRAM storage which allows for long exposure shots. An automatic dark signal contribution mitigation scheme has been implemented to enhance the visual quality in dark areas. Global analog reference to the pixels is dynamically distributed to allow for low-power, fast, and precise operation.

In order to increase fill factor, resolution and light sensing performance, a possible evolution of the system can be an implementation as part of a 3D integrated system. The idea in shown in Fig. 9 where the dies are connected Through Silicon Vias (TSV). The die on top (Tier 0) could contain the photodiode, which allows to use a die of a Back Side Illuminated technology to improve sensing capabilities and improve fill factor to near 100% also permitting to reduce the pixel size. The die in the middle (Tier 1) could contain the rest of the circuitry of the pixel, which allows to use a technology enhanced for mixed circuitry. The die at the bottom (Tier 2) could contain the circuit to post-process over the final tone mapped image, which allows to use a die with a technology enhanced for digital circuitry that usually offer very high integration density. The levels per bin calculations and the LUT could be contained in the last tier, as it could contain only digital processing, or it can be included laterally in tier 1.

## ACKNOWLEDGMENTS

This work is partially funded by TEC2009-11812, CENIT ADAPTA, ONR Grant N000141110312, and FEDER 2007-2013.

## REFERENCES

1. S. Vargas-Sierra, G. Liñán Cembrano, and A. Rodríguez-Vázquez, "High-dynamic range tone-mapping algorithm for focal plane processors," in *SPIE Microtechnologies*, 2011., April 2011.
2. A. Belenky, A. Fish, A. Spivak, and O. Yadid-Pecht, "Global shutter cmos image sensor with wide dynamic range," *Circuits and Systems II: Express Briefs, IEEE Transactions on* **54**, pp. 1032–1036, dec. 2007.
3. M. Mase, S. Kawahito, M. Sasaki, Y. Wakamori, and M. Furuta, "A wide dynamic range cmos image sensor with multiple exposure-time signal outputs and 12-bit column-parallel cyclic a/d converters," *Solid-State Circuits, IEEE Journal of* **40**, pp. 2787–2795, dec. 2005.
4. H.-Y. Cheng, B. Choubey, and S. Collins, "An integrating wide dynamic-range image sensor with a logarithmic response," *Electron Devices, IEEE Transactions on* **56**, pp. 2423–2428, nov. 2009.
5. A. Spivak, A. Belenky, A. Fish, and O. Yadid-Pecht, "Wide-dynamic-range cmos image sensors - comparative performance analysis," *Electron Devices, IEEE Transactions on* **56**, pp. 2446–2461, nov. 2009.
6. E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*, Elsevier / Morgan Kaufmann, 2006.
7. Sony, "Cybershot dsc-w80," <http://news.sel.sony.com/assets/Cyber-shot.2007/specs/DSC-W80.pdf>.
8. Apple, "Iphone 4 user guide," [http://manuals.info.apple.com/en\\_US/iphone\\_user\\_guide.pdf](http://manuals.info.apple.com/en_US/iphone_user_guide.pdf).
9. Photonfocus, "Mv-d752e-40-u2-12," <http://www.photonfocus.com/html/eng/products/products.php?prodId=55>.